NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

RELATING THE DISTRIBUTIONAL CHARACTER
OF NUMERICAL MODEL OUTPUT PARAMETERS
TO THE OCCURRENCE OF FOG
OVER THE NORTH ATLANTIC OCEAN

by

Oliver J. Muldoon

June 1986

Thesis Advisor: R. J. Renard
Thesis Co-advisor: P. R. Lowe

Approved for public release; distribution unlimited

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

1b RESTRICTIVE MARKINGS:

2a SECURITY CLASSIFICATION AUTHORITY:

2b DECLASSIFICATION / DOWNGRADING SCHEDULE:

3 DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution unlimited

4 PERFORMING ORGANIZATION REPORT NUMBER(S):

5 MONITORING ORGANIZATION REPORT NUMBER(S):

6a. NAME OF PERFORMING ORGANIZATION: Naval Postgraduate School

6b OFFICE SYMBOL (If applicable): 63

7a NAME OF MONITORING ORGANIZATION: Naval Postgraduate School

6c ADDRESS (City, State, and ZIP Code):

7b ADDRESS (City, State, and ZIP Code): Monterey, California 93943-5100

8a NAME OF FUNDING / SPONSORING ORGANIZATION:

8b OFFICE SYMBOL (If applicable):

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER:

8c ADDRESS (City, State, and ZIP Code):

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification): RELATING THE DISTRIBUTIONAL CHARACTER OF NUMERICAL MODEL OUTPUT PARAMETERS TO THE OCCURRENCE OF FOG OVER THE NORTH ATLANTIC OCEAN

12 PERSONAL AUTHOR(S): Muldoon, Oliver J.

13a TYPE OF REPORT: Master's Thesis

13b TIME COVERED: FROM ________ TO ______

14 DATE OF REPORT (Year, Month, Day): 1986 June

15 PAGE COUNT: 58

16 SUPPLEMENTARY NOTATION:

17 COSATI CODES

FIELD	GROUP	SUB-GROUP

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Marine fog; distribution of model predictors; Bayes' Law; Beta, Gamma, Normal distributions of parameters associated with Fog/No Fog

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

This report describes an investigation into the statistical distributional character of six model output parameters from Fleet Numerical Oceanographic Center's Navy Operational Global Atmospheric Prediction System, as a function of the occurrence of Fog and No Fog for a climatologically-homogenous area of the North Atlantic Ocean in the summer season. Beta, Normal and Gamma distributions were fitted to these parameters and forecasts, using various categorical scoring systems (Threat Score, Percentage Correct, False Alarm, Forecast Reliability and Power of Detection), as well as a probabilistic scoring system (Penalty-Reward Score). The forecast results were examined for significant differences using an Anova analysis. It is confirmed that predictor populations whose underlying distributions are of an exponential form are much better

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION: unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL: Robert J. Renard

22b TELEPHONE (Include Area Code): 408-646-2516

22c OFFICE SYMBOL: 63Rd

19. continued

represented by a Beta or Gamma distribution than by a Normal. For predictors whose distributions are quasi-bell shaped, it is indicated that the Beta Distribution cna be generally used as a proxy for the Normal, as can the Gamma also. However, in some cases the Normal distribution results in better forcase scores, and the decision on use of a proxy would depend on which of the scores is to be emphasized.

Approved for public release; distribution is unlimited.

Relating the Distributional Character of Numerical Model Output Parameters to the Occurrence of Fog over the North Atlantic Ocean

by

Oliver J. Muldoon
Captain, United States Air Force
B.A., University College, Galway, Ireland, 1970

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

NAVAL POSTGRADUATE SCHOOL
June 1986

ABSTRACT

This report describes an investigation into the statistical distributional of six model output parameters from Fleet Numerical Oceanographic Center's Navy Operational Global Atmospheric Prediction System, as a function of the occurrence of Fog and No Fog for a climatologically-homogeneous area of the North Atlantic Ocean in the summer season. Beta, Normal and Gamma distributions were fitted to these parameters and forecasts of Fog and No Fog were made using Bayes' Law. Intercomparisons were made of these forecasts, using various categorical scoring systems (Threat Score, Percentage Correct, False Alarm, Forecast Reliability and Power of Detection), as well as a probabilistic scoring system (Penalty-Reward Score). The forecast results were examined for significant differences using an Anova analysis. It is confirmed that predictor populations whose underlying distributions are of an exponential form are much better represented by a Beta or Gamma distribution than by a Normal. For predictors whose distributions are roughly bell-shaped, it is indicated that the Beta distribution can be generally used as a proxy for the Normal, as can the Gamma also. However, in some cases the Normal distribution results in better forecast scores, and the decision on use of a proxy would depend on which of the scores is to be emphasized.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I'm very grateful to the following people at the Naval Postgraduate School on whom I relied for help and encouragement: Professor Robert Renard, my thesis advisor; Mr Paul Lowe, my co-advisor; Mr Robert Hale, my day-to-day source of guidance, and my fellow Galwegian, Dr. Iognáid Ó Muirceartaigh, for his statistical advice. In addition, I'm thankful to my wife, Susan, and my children Frank, Peter, Laura, Fióna and Shena, for their love, patience and understanding over the past two years. I'm also grateful to the United States Air Force for giving me the opportunity to attend graduate school. Finally, my thanks to my heavenly Producer, without Whom this effort would've have been impossible.

I. *INTRODUCTION AND OBJECTIVES*

A. GENERAL

Marine fog is a hazard to shipping and to low-level flying over the open ocean and coastal waters. Wheeler and Leipper (1974) cited human and other losses to the United States Navy due to poor visibility caused by fog. On the other hand, marine fog can camouflage the location and motion of surface shipping. This works both ways; it helps protect friendly forces from discovery but makes it more difficult to seek out and destroy enemy forces. Since the Strategic Air Command is responsible for interdicting enemy sea power through air operations and for conducting antisubmarine warfare and aerial minelaying operations, forecasting marine fog is of great importance to the U.S. Air Force as well as to the U.S. Navy.

B. ROLE OF THE NAVAL POSTGRADUATE SCHOOL

In recent years, the Department of Meteorology, Naval Postgraduate School (NPS) has been studying the climatology of marine fog and marine visibility (Renard, Englebretson and Daughenbaugh, 1975; Willms, 1975; Renard, 1976). However, the data network is not widespread enough over the oceans to generate sufficient information to analyze the initial visibility/fog conditions as a prerequisite for forecasting marine fog on a day-to-day basis.

Because of the difficulties of forecasting fog directly, NPS researchers began to use Model Output Statistics (MOS) to estimate marine visibility (and implicitly marine fog). This effort has been concentrated mostly on areas of the North Pacific Ocean (Koziara, Renard and Thompson, 1983; Renard and Thompson, 1984). Karl (1984), Diunizio (1984) and Elias (1985) worked with a number of climatologically homogeneous areas of the northwest North Atlantic Ocean. These researchers tested various MOS prediction schemes, using predictors from the Fleet Numerical Oceanography Center's (FNOC) Navy Operational Global Atmospheric Prediction System (NOGAPS). Fatjo (1986) tested these same schemes on controlled (simulated) data sets.

C. RESULTS TO DATE

Results so far suggest the problem of forecasting marine visibility is even more intractable than previously supposed. Because of the difficulties of forecasting visibility

using more than two categories, Karl and Diunizio recommended switching to a two-category visibility forecasting scheme. Also, Diunizio recommended including derived predictors as well as direct model predictors; a derived predictor is a mathematical combination of two or more direct model predictors. Fatjo found that, in general, a good predictor can overcome gross data defects, suggesting it would be more profitable to improve the quality of a few key predictors than undertake a costly and massive upgrade to the data network. Fatjo also showed that the degree of statistical separation between data sets is of the utmost importance in statistical forecasting using fitted distributions.

As a result, Lowe[1] recommended a study of the statistical attributes of the predictor data sets. He also expressed great unease at assuming the underlying distribution for all predictors to be Normal, as has hitherto been the case. While seeing the attraction of assuming a common distribution for each predictor (rather than having to fit each one individually), Lowe questioned whether there might be a more generally applicable distribution than the Normal.

D. PURPOSE OF THIS STUDY

Ultimately, it became clear that learning more about the statistical distribution of model output predictors vis-a-vis the occurrence of Fog and No Fog is a necessary ingredient in a MOS forecasting scheme for marine fog. In particular, it was felt useful to know for which predictors the Beta distribution may be safely substituted for the Normal distribution.

Accordingly, this study presents the results of an investigation into the distributional character of certain model output parameters which could be regarded as potential marine fog predictors. In the process, three different distributions (Beta, Normal and Gamma) are compared, although the emphasis is on comparing the Beta and Normal distributions.

E. SUMMARY OF STEPS TO BE TAKEN

First, determine the most likely NOGAPS predictors which might serve as fog predictors.

---

[1] P. R. Lowe is a senior scientist at the Naval Environmental Prediction Research Facility, Monterey, CA. He is also Co-Advisor to this study.

Second, establish a working data base, consisting of observed data matched with NOGAPS predictors.

Third, compute the Beta, Normal and Gamma distributional forms of the predictors, along with the relevant statistical parameters of these distributions.

Fourth, graphically fit each predictor population to the Beta, Normal and Gamma distributions.

Fifth, for each distribution, apply Bayes' Law of Inverse Probability to diagnose (i.e."predict") the occurrence of Fog or No Fog.

Sixth, determine for which predictors the Beta distribution is competitive with the Normal.

Seventh, compare the Gamma distribution with the Beta and Normal distributions.

II. *MODEL OUTPUT STATISTICS*

A. GENERAL

A Model Output Statistic (MOS) is a statistically-developed method which forecasts a weather element of interest as a function of forecast variables available from a numerical weather prediction model. While some common meteorological variables are directly forecasted by the numerical model, like pressure and temperature, there are some important exceptions, such as fog and visibility. The Navy's MOS program is being developed at the Naval Environmental Prediction Research Facility (NEPRF) to fill in that gap by producing forecasts of the elements not forecasted by the numerical model. The MOS procedure is also used to refine and tailor numerical model forecasts to account for model errors and sub-synoptic scale influences.

B. MODEL OUTPUT PREDICTORS

Meteorological variables directly forecasted by the numerical model are usually called model output predictors (MOP). A MOP is the model's "best guess" of the value of that variable at a particular point in space and time.

The set of predictor values for a variety of meteorological variables may be imagined as an array of numbers, defined for that point in space and time. Considering the number of geographical points and forecast periods over which the numerical model produces arrays at a given time, there's a massive volume of data involved. Portions of these data are usually archived at weather centers.

In theory, if each numerical-model predictor array were perfectly accurate, it would predict the exact state of the atmosphere at a given time and locus. A statistical analysis of both the predictors and observations for that point in space, over a suitably long period of record, would be expected to show at least one predictor having a markedly different distribution of values between two events, such as Fog and No Fog. Thus, it's concluded that this predictor (or group of predictors) "makes the difference" between whether Fog is present or absent.

As a crude example, assume this predictor to be the relative humidity (RH). It might be that Fog occurs always and only when RH is, say, 90% or more, and, conversely, No Fog occurs always and only when the RH is 89% or less. The search for a perfect predictor of Fog, for that point in space, would be over.

C. MODEL OUTPUT PREDICTORS AND OBSERVED ELEMENTS

Numerical model predictions are usually made and disseminated twice daily. Glahn and Lowry (1972) recognized that these model output predictors, together with the corresponding archived weather observations, collectively represent a wealth of valuable information. If a certain predictor value (or range of values) could be shown to be consistently related to the value (or range of values) of a certain observed variable, an association between the two could be established which, hopefully, would hold true in the future.

As a result of Glahn's and Lowry's work, the Technique Development Laboratory of the National Weather Service began working on prediction equations that would establish statistical relationships between model output predictors and various weather elements of interest, known as predictands. The statistical relationships are determined by multiple linear regression[2] (Glahn, 1983).

D. SOME LIMITATIONS TO MOS FORECASTING

The basic assumption in MOS forecasting is that some relationship between the predictor and predictand, established from historical data, is valid under similar circumstances in the future. Therefore, the quality of the predictand data is crucial. These data consist of weather observations taken under widely varying circumstances. They play two roles: they serve as initial data for the numerical model, and they verify (where feasible) the accuracy of the predictions (i.e. MOS predictors).

Unfortunately, the raw observations initializing the numerical model don't come close enough to capturing the true state of the atmosphere. For forecasting over the ocean, there aren't enough unique observations regularly positioned over an area of interest, leaving large areas with data gaps. Except for some scattered weather buoys, observation platforms (i.e. ships) report from different locations as they travel, often at irregular intervals. Observers' skills vary widely, from those of a trained meteorologist to an ordinary crewman with minimal experience. Of special importance to this study, horizontal visibility reports at sea are hampered by lack of visibility markers such as are usually present on land. The weather transmission code limits the degree of detail that a reported observation may include. The transmission network from ship to weather center can delay, garble and lose observation data.

---

[2] a greatly modified form of Regression Estimation of Event Probability (REEP), according to P. R. Lowe of the Naval Environmental Prediction Research Facility.

The numerical model is a discrete-point representation of a spatial-temporal continuum. Consequently, it has to cut mathematical corners in order to reduce the complexity of the calculations and to meet forecast deadlines. Also, there still exists an imperfect and incomplete understanding of the complex dynamic processes of the atmosphere, especially on the smaller scales of time and space.

Even if the observation limitations were corrected and the numerical model was made more sophisticated, and thus more accurate in its predictions of atmospheric variables, the basic problem of an imperfect match of predictors and predictands still remains. The relationships between the two are generally very complex and highly nonlinear, and MOS prediction equations can only attempt to capture the complex feedback mechanisms between the different atmospheric variables.

E. NOGAPS MOS DATA

In 1983, the U.S. Navy began developing a MOS computer program to forecast horizontal visibility at sea, using the FNOC NOGAPS model. NOGAPS produces predictor values for a variety of meteorological variables at global grid-points, spaced at intervals of approximately 2.5^0 latitude and longitude. These are the raw material from which MOS forecasts are produced. The data have been archived at FNOC and were made available for this study.

III. *STATISTICAL OVERVIEW*

A. THE CONCEPT OF SEPARATION

One of the attractions of statistics is its ability to extract valuable, hidden information from a mass of raw data. For example, suppose there are n observations each of Fog and No Fog for an ocean location. A plot of the occurrence of Fog versus Predictor A might result in a distribution like Fig. 3.1a, with values of 290 occurring most frequently, and with other values less frequently on each side of the mode. A similar plot for No Fog might look like Fig. 3.1b, with a peak at 297. When these two figures are combined into one (Fig. 3.1c), the relative frequencies can be readily compared.

From Fig. 3.1c, it follows that a predictor value of T_1 has a P_1 likelihood of being associated with Fog, and a Q_1 likelihood with No Fog. Since P_1 and Q_1 are about equal, it can be seen that a T_1 value of Predictor A has about the same chance of happening in either case and is thus of little associative use. In other words, given a value of T_1, valid for a particular point in time and space, and associating (i.e. forecasting) T_1 at that time with, say, Fog, a correct forecast would be expected slightly more than half the time, given repeated forecasts over the long term. This is because P_1 is only slightly greater than Q_1. Since this is little better than tossing a coin, this value of T is not a very skillful predictor.

However, a value of T_2, with likelihoods P_2 and Q_2 of Fog and No Fog, respectively, would be an excellent associative tool; a forecast of the event with the higher likelihood would be expected to be right much more than half the time. This is because Q_2 is much greater than P_2.

In this example, since most values of Predictor A exhibit a big difference between their respective P and Q values, this translates into the ability to make a correct forecast most of the time.

Clearly, the attribute of the data that enables us to make such a confident forecast is an important one; it might be described as the ability to discriminate between different events with an acceptable degree of confidence. It will be referred to here as separation.

Separation is perhaps best illustrated when it's absent altogether. If it was found that the distribution of Predictor A was exactly the same for both Fog and No Fog, it

would be concluded that this predictor had the same range and frequency of values in both cases...that it couldn't distinguish between the two events. Hence, it would show zero skill as a predictor. Fig. 3.1d is an example where the frequency distributions are so close as to be almost indistinguishable.

Separation may be achieved in a number of different ways. The type of distribution itself, and the mean and standard deviation (hereafter referred to as sigma), all play an important role. A quantitative measure of separation will be defined later in this study; meanwhile, the following examples illustrate some of the types of separation that are possible:

1. Given a common distributional form, roughly bell-shaped with common means, the respective sigmas determine the separation. In Figs. 3.2a and 3.2b, the sigmas are small, but the bigger difference between them in Fig. 3.2b makes for better separation than in Fig. 3.2a. In Figs. 3.2c and 3.2d, the sigmas are larger (Note: Figs. 3.3 and 3.4 are scaled differently); again, the separation is better in the case of the larger sigma difference.
2. Given a common distributional form, roughly bell-shaped but with different means, the picture becomes more complicated, with the two means and two sigmas influencing the separation. Fig. 3.3a illustrates the statistician's dream of two populations that are mutually exclusive. The sigmas are both small compared to the intermean distance. Fig. 3.3b depicts the intermean distance as large compared to one of the sigmas. The common area beneath the two curves is relatively small, which, of course, is one of the ways separation can occur. Fig. 3.3c shows a small intermean distance compared to both sigmas, meaning little separation. Finally, Fig. 3.3d's intermean distance is small compared to one of the sigmas; some of the separation is attributable to the relatively large area of mutually exclusive events at the tails of the flatter curve, while the rest of the separation is due to the sharp difference in the relative frequencies in the vicinity of the means of each population. This case is merely a variation of Figs. 3.2b and 3.2d, where the means are equal.
3. Given one or more non-bell-shaped distributions, separation may also be attainable. Fig. 3.4a shows a normal and exponential distribution, but with a common mean and sigma. Despite that, there's excellent separation in the area of the highest exponential density and moderate separation toward the right-hand side of the graph. The forecasting problem is acute only at and near the intersection of the curves. Fig. 3.4b depicts two exponential distributions; again, the "forecasting" task is dubious only over a narrow range of data values, which is most desirable. Finally, Figs 3.4c and 3.4d, which look familiar from the above examples, are actually manifestations of the Beta distribution. Fig. 3.4c is of the general form of Fig. 3.4a, while Fig. 3.4d looks just like the distributions in the example of the Predictor A in Fig. 3.1c. Incidentally, there's excellent separation in both cases.

The chameleonic property of the Beta distribution, illustrated in the last example, is part of the subject of this thesis, and will be looked at again later.

B. QUANTIFYING SEPARATION

From the above discussion, it can be seen that the statistical separation of two data sets is a function of the respective means and sigmas, themselves functions of the distribution of the data. To quantify separation, a number called the Signal-to-Noise

Ratio (S/N) has been developed by Lowe[3] where

$$S/N = (\delta\mu)^2(N_1 + N_2 - 2)/(N_1\sigma_1^2 + N_2\sigma_2^2).$$

S/N is a measure of the intermean distance $\delta\mu$, modified by the sizes (N_1, N_2) and standard deviations (σ_1, σ_2) of the respective populations.

However, S/N doesn't tell the whole story, as a glance at Fig. 3.4a shows. Even though the means and sigmas are equal, there's still significant separation between the data sets, and forecasting one or the other of the two events has good prospects of success, especially for lower values of the data. *Hence, an important caveat to the use of the S/N is that the data sets themselves must be examined to determine their characteristic distributions.* If the characteristic distributions are dissimilar, this fact itself often overrides whatever indications the S/N might give.

C. REAL DATA

Of course, real-world data isn't as clear-cut as the above examples, and separation is sometimes an elusive goal. For one thing, such data may be only roughly fitted to its "characteristic" distribution, with a poor "goodness-of-fit" at times. Also, real data are often contaminated; for example, outlying values which may be physically unrealistic may nevertheless creep in due to data processing problems. In short, the real world data are considerably more messy, and occasionally of little forecasting use.

In the case of output from a numerical weather prediction model, such as MOS predictors, even the "best" predictors can be expected to have a fair degree of imprecision. After all, these are only as good as the raw input data, and subject to mathematical and physical simplifications. By finding out which predictor works best for a given meteorological phenomenon like Fog, including learning as much as possible about that predictor's statistical distribution, resources can be concentrated on improving the model's ability to forecast with that particular predictor.

In this study, 59 NOGAPS predictors were available, a number of which were combined to form derived predictors. These are listed in Appendix C. Since it was necessary to find some way to reduce this number to a more manageable one, *a mix of statistical and meteorological insights* was used to arrive at a short list of candidate variables.

---

[3] P. R. Lowe, Naval Environmental Prediction Research Facility, Monterey, CA.

D. PREDICTOR SELECTION

Statistical insights led to computing the S/N for each predictor, and rank ordering them by S/N value. Concurrently, histograms of each predictor were plotted to estimate and compare the overall shape of the predictor's distributions for Fog and No Fog. Those predictors whose S/N ratios were less than 0.5 and with broadly similar distributions for both populations were eliminated from further consideration.

Candidate predictors were then examined from a meteorological perspective to see if they made sense physically. As expected, those predictors at the top of the list were related in some way to the marine atmospheric boundary layer or air-sea interface. However, it is surprising that the derived predictors Surface Relative Humidity, Surface Air Temperature Advection and Sea-surface Temperature Advection showed no promise as statistical predictors. These advective quantities were computed using numerical model wind and temperature data.

Time constraints restricted detailed examination to the following six predictors:

1. *Surface Moisture Flux (SMF)*

S/N: 1.006. This is equivalent to Evaporative Moisture Flux, which Koziara et al (1983) found to be the best parameter for predicting marine fog over the North Pacific Ocean. A downward flux results when a moist airmass is cooled to saturation at the sea surface, setting up a favorable condition for fog formation and maintenance.

2. T_{925} *- SST (TDF)*

S/N: 1.510. This is defined as the difference between the air temperature at 925mb and the sea-surface temperature, which Diunizio (1984) recommended as a prospective predictor. The 925mb temperature gives a better S/N than does the surface air temperature. Marine fog is usually associated with a negative difference between these quantities. The conjunction of relatively cold air and warm water over the Gulf Stream in the location examined in this study would be expected to amplify this difference.

3. *Entrainment (ENT)*

S/N: 0.777. The degree of turbulent mixing in the marine boundary layer should be reflected in this predictor. Barker (1975) states that entrainment from the inversion layer into the boundary layer is an important source of heat and drier air, effectively acting to retard fog formation.

4. *Long-wave Radiation (LWR)*

S/N: 0.732. Fog tends to trap outgoing long-wave radiation from the surface, and reradiate some of it back down. Thus, the Fog and No Fog regimes might be expected to show differences in LWR profiles, with the incidence of Fog negatively correlated with LWR. Renard and Thompson (1984) found infrared extinction parameters to be useful in their work on visibility over the North Pacific Ocean.

5. *Sensible Heat Flux (SHF)*

S/N: 0.816. Mack et al (1983) link a downward heat flux with marine fog formation. A downward flux reflects the movement of warmer low-level air over a colder ocean surface, a condition necessary for advection fog over the ocean. A link was also shown between a slight upward heat flux with the maintenance of fog poleward of the area of maximum sea-surface temperature gradient.

6. *Stratus Frequency (STF)*

S/N: 0.626. Pilie et al (1979) linked the occurrence of stratus with coastal fog off California. This is not surprising since fog is surface-based stratus, and the frequency of the two would be expected to be related.

IV. *DATA*

A. AREA

The area of study was confined to a region in the North Atlantic Ocean, one of a number of climatologically homogeneous regions.[4] This area, off the coast of Newfoundland, has a relatively high occurrence of marine fog compared to the other regions in the North Atlantic Ocean. (Renard, 1980). This region is identified in Fig. 4.1.

B. TIME

Data from the months of June, July and August, 1984 and 1985, were meshed into a single data set. The only synoptic ship reports used were those at 1200 GMT, since this is a daylight hour over the region. It was felt that synoptic data at 0000 GMT, under fading light conditions at best, would not be as reliable for the detection of fog.

C. RAW VERIFICATION DATA SET

A "raw" data set was compiled by the Naval Oceanography Center Detachment, co-located with the National Climatic Data Center (NCDC), in Asheville, NC. The data set consists of all ship synoptic observations for this study region and period on file at NCDC.

Each report has been graded by NCDC for accuracy and consistency, from the temporal, spatial and meteorological perspectives. Flags have been inserted in the data as an indicator of reliability and to alert the user to questionable reports.

D. REFINED VERIFICATION DATA SET

To refine the raw verification data set, certain deletions became necessary. Reports whose geographical location had been questioned by NCDC were discarded. Reports which provided no reliable evidence of either the presence or absence of fog were deleted. An example would be a buoy, which might report only a sea-surface temperature. Also deleted was one of multiple observations from the same location and time. The observation retained was the one deemed most reliable by NCDC. In

[4] as determined by P. R. Lowe of the Naval Environmental Prediction Research Facility, Monterey, CA.

addition, reports close to shore were eliminated; these are susceptible to contamination from land-based predictor values during subsequent interpolation of such values to near-shore ship observation locations.

Once these deletions were made, the data set was divided into two parts corresponding to Fog/No Fog. The criteria for this division are listed in Appendix B.

E. MODEL OUTPUT PREDICTORS

While the verification data set consists of surface observations at scattered points over the area in question, NOGAPS predictions are made for grid points. The following steps were taken to prepare these data for use:

1. *Interpolation to Observation Points*

The predictor values were interpolated from grid points to ship observation points, using a bilinear interpolation technique. Thus, the predictor data set was reduced in length so as to be equal to the length of the accepted surface observations data set.

2. *Predictor Subsets*

The data set was divided into subsets corresponding to each predictor; the size of each predictor subset depends on the number of missing predictor values; if there weren't any, the subset length equals that of the overall predictor data set.

3. *Fog/No Fog Subsets*

Each predictor subset was further divided in two, based on whether the associated verification observation indicated Fog or No Fog.

V. *MODUS OPERANDI*

A. PREPARATORY

Once the predictor data were divided into subsets corresponding to the Fog and No Fog populations of each one, as described earlier, the S/N test was conducted on each predictor data set. A short list of candidate predictors was then drawn up, forming the basis of all further work in this study.

B. COMPUTER PROGRAM

For each candidate predictor, a computer program program first randomly splits in half the two candidate predictor populations, Fog and No Fog, forming a training and a testing set. Once the data are split, the mean (μ) and sigma (σ) of the Fog and No Fog training sets are used to compute bounds to the range of the data, A and B, defined as follows:

$$A = \mathrm{MIN}\,(\mu_1 - 3\sigma_1,\ \mu_2 - 3\sigma_2), \text{ and}$$

$$B = \mathrm{MAX}\,(\mu_1 + 3\sigma_1,\ \mu_2 + 3\sigma_2).$$

Data values outside these bounds were discarded in order to eliminate the undue influence that such outlying values exert when trying to fit a distribution to the remaining values. Thus, the maximum and minimum values of the remaining data became the range over which the distributions were fitted.

C. TRAINING SET

The training set was used to generate statistics that are assumed to characterize the population as a whole. This assumption was based on visual comparisons of the empirical distributions, as discussed below.

1. *Empirical Distribution of Training Data set*

Both the Fog and No Fog training sets were inserted into Grafstat, an interactive statistics package, which is described in more detail in Appendix A. In Grafstat, an empirical plot of each data set was made on the same graph. This served as a check on the randomness of the splitting procedure, since one would expect to see the same overall pattern as with the entire data set. This also gave an idea of the overall shape of the distribution. Finally, the plot showed at a glance the separation of the two

populations (or lack thereof). Empirical plots of the training sets of the six predictors are shown in Fig. 4.2.

2. *Fitting Distributions*

While in Grafstat, the training populations were fitted to the Beta, Gamma and Normal distributions, using the Method of Moments procedure. This method was chosen over the Maximum Likelihood method due to the difficulty of adapting the latter to a Fortran computer program. These fits are shown in Appendix D.

3. *Using Distribution Parameters*

The main computer program used in this study computed the statistical parameters of the Beta, Gamma and Normal distributions. The accuracy of these was checked against corresponding values generated by Grafstat.

4. *Output from the Training Sets*

The statistics gleaned from the training set would be applied later to samples drawn from the testing set, in the hope of distinguishing the Fog from the No Fog cases in the latter. These statistics are the parameters for the Beta, Normal and Gamma distributions, α, β, μ, σ, λ, and n respectively.

D. TESTING SET

Ten samples were randomly drawn from each population of the testing set, 10% in length. This preserved the same relative frequency between Fog and No Fog cases as in the original data sets. For each value in each sample, the goal was to determine the probability of its belonging to one or other of the populations. Since the two populations were of different sizes, with different prior probabilities, recourse was made to Bayes' Law.

1. *Bayes' Law*

The forecast method involved applying Bayes' Law to each of the sample predictor values. This law takes into account the prior, unconditional probabilities of each event. For the purposes of this study, the prior probabilities were defined by the relative sizes of the two populations (approximately 65% and 35% for No Fog and Fog respectively). Bayes' Law for the conditional probability of the event Fog, given the occurrence of a predictor value A, may be stated as follows:

$$P(\mathrm{Fog})|A = \frac{P(\mathrm{Fog}) \times f(A|\mathrm{Fog})}{P(\mathrm{Fog}) \times f(A|\mathrm{Fog}) + P(\mathrm{No\ Fog}) \times f(A|\mathrm{No\ Fog})},$$

where P(Fog), P(No Fog) are the unconditional (prior) probabilities, based on the relative sizes of the two training populations; and f(A|Fog), f(A|No Fog) are the class conditional likelihoods that a value A will be associated with Fog or No Fog respectively. These likelihoods are computed using the statistical parameters obtained from the training set, applied to the three distributions being examined. An analogous equation may be written for P(No Fog|A), the conditional probability of No Fog.

2. *Scoring the Forecasts*

Whenever Bayes' Law computed a probability of $\geq 50\%$ that a predictor value, known to be from a certain population, indeed came from that population, it was considered a hit. A tally was kept of the number of hits and misses per event, sample and distribution. Contingency tables were generated for each sample and a number of different skill scores were computed. These scores are defined in Chapter VI.

VI. *SCORING AND TESTING FOR SIGNIFICANCE*

A. GENERAL

The overall goal was to determine if the Beta distribution could serve as a proxy for the normal distribution for certain meteorological predictors produced by the NOGAPS model. In the process, the Gamma distribution was also tested. Using the statistical parameters of each distribution, forecasts were made and the results compared for significant differences. For each predictor, three two-way comparisons were made...Beta-Normal, Beta-Gamma and Normal-Gamma.

B. SCORES COMPUTED

Each sample drawn from a testing set represents one set of forecasts. For each forecast, five percentage scores were derived from contingency tables. These scores are defined as follows (an incorrect forecast of Fog means Fog was forecasted but not observed):

1. Threat Score (TS): Number of correct forecasts of Fog divided by the sum of all observations of Fog and all incorrect forecasts of Fog.
2. Percentage Correct (PC): Number of correct Fog and No Fog forecasts divided by all forecasts.
3. Power of Detection (PD): Number of correct forecasts of Fog divided by all observations of Fog.
4. Forecast Reliability (FR): Number of correct forecasts of Fog divided by all forecasts of Fog.
5. False Alarm Rate (FA): Number of incorrect forecasts of Fog divided by all observations of No Fog.

These scores were also averaged over ten repetitive sample runs for an overall score. Except for the False Alarm score (FA), the higher the score the better the forecast.

C. PENALTY-REWARD SCORE

In addition to the contingency table scores, each forecast was scored using the Penalty-Reward (PR) score.[5] This score measures the skill of the probabilistic forecasts of two-category events, such as Fog/No Fog.

---

[5]devised by P. R. Lowe of the Naval Environmental Prediction Research Facility, Monterey, CA

For this study, the PR score is defined for the dichotomous case of Fog/No Fog, where P(No Fog) is greater than P(Fog). A separate PR score is computed for each individual forecast made, and an overall PR score is calculated as the mean of the individual scores. *Only the overall scores are shown in this study.*

The PR score is computed as follows:

For $P(Fog|A) < P(Fog)$, $PR = (I_2 - I_1(X-1))(1-Y)^2$

For $P(Fog|A) > P(Fog)$, $PR = (I_1(X-1) - I_2)((P(Fog|A) - P(Fog))/(1-P(Fog)))^2$

where

1. P(Fog|A), P(Fog) are as defined for Bayes' Law
2. $I_2 = 0$ if Fog occurs, $= 1$ if No Fog occurs
3. $I_1 = 1$ if Fog occurs, $= 0$ if No Fog occurs
4. $X = 1/P(Fog)$
5. $Y = P(Fog|A)/P(Fog)$

D. TESTING FOR SIGNIFICANCE

A one-way Analysis of Variance (Anova) procedure, using ten samples, is used to test for significant differences between the forecast scores obtained by pairs of distributions. The results for each two-way Anova comparison are given in Appendix E, where AnovaBN, AnovaBG and AnovaGN refer to comparisons between Beta and Normal, Beta and Gamma, and Gamma and Normal respectively. The level of significance was set at 0.05. Values less than this number indicate a significant statistical difference exists between the two scores being compared.

E. JUDGING THE RESULTS

Since the hypothesis to be validated is that the Beta distribution may be used as a proxy for the Normal distribution, this is equivalent to seeking a significant, negative difference between these two distributions. There are five possible significance profiles that could occur:

1. Beta could be significantly better than Normal.
2. Beta could be insignificantly better than Normal.
3. Beta could be exactly the same as the Normal.
4. Beta could be insignificantly worse than Normal.

5. Beta could be significantly worse than Normal.

Of this list, only item 5 would nullify the basic hypothesis. Analogous comparisons between Beta and Gamma, and Gamma and Normal were also made.

VII. *RESULTS*

A. GENERAL

For each predictor examined, three histograms of each population (Fog and No Fog) were generated using Grafstat. Pairs of the three distributions being examined were fitted to each histogram, allowing three two-way visual comparisons of the ability of these distributions to fit the data. In generating these histograms, the data were standardized between 0 and 1 for each population separately, since these are the limits of a Beta distribution.

Tables 1 and 2 show the results of the forecasting procedure for each of the three distributions used; these are contained in the first three lines of each predictor segment. The next three lines show the Anova probability values (P-values); these are for three one-way analyses of different pairs of distributions, and are identified by the first initial of each member of the particular pair being examined. A P-value less than 0.05 was taken to indicate significant differences between scores.

The last line ranks the three distributions in descending order of forecasting skill, using the first initial of each one. Where the Anova results show no significant difference between all three, the initials are not separated, i.e. BNG. Significant differences are indicated by a dash between the distributions. For example, B-N-G indicates a significant difference between all three. BN-G indicates that Gamma is significantly worse than both Beta and Normal, while the latter are not significantly different from each other. A few cases occurred where the first and third ranked distributions were significantly different from each other but not from the second ranked one; this is indicated in the table.

In making the forecasts, time and programming constraints dictated that the same distribution be used on the Fog and No Fog populations, as opposed to mixing the distributions by using the "best fit" ones for each population.

B. SURFACE MOISTURE FLUX (SMF)

The empirical distributions (Fig. 4.2) are somewhat bell-shaped, with a fair degree of visual separation between the populations. Separation is enhanced since the No Fog population is clearly more dispersed than the Fog population.

Looking at the histograms with the distribution fits (Fig. 7.1), the Gamma distribution fits both populations very well. For the Fog population, Gamma appears to capture the peak of the data a little better than the Beta, as well as to reflect the skewness more accurately. The Normal is clearly less able to fit these data, and, of course, is unable to reflect any skewness at all. For the No Fog case, there's little difference between Beta and Normal. However, the Gamma distribution does better than either of the others both in showing the overall shape of the data and in capturing the skewed peak. Visually, then, the Gamma is the best fit, followed by the Beta and Normal in that order, for both populations.

The forecasting scores show TS values between 0.426 and 0.492, comparable with those found in the North Pacific Ocean experiments referenced in Chapter I. The TS value is perhaps the most important score meteorologically, since it measures the ability of a predictor to forecast "threatening" events, such as fog. For TS, there is no significant difference between Beta and Normal, both of which are significantly better than Gamma. This is surprising, since the Gamma is the better fit visually.

The PC score shows no significant difference between the distributions, with Beta slightly better than Normal. However, significant differences are present in each of the other scores, as seen from Table 1.

Only in the PD score is the Normal significantly better than the Beta. For the others, either there is no significant difference between these two or Beta is significantly better than Normal. The only reflection in the forecast scores of Gamma's visually superior fit to both populations is its showing in FR and FA, where it is significantly better than the others.

C. T_{925} - SST (TDF)

The empirical distributions (Fig. 4.2) are again bell-shaped, with a fair degree of visual separation between the populations.

Looking at the histograms with the distribution fits (Fig. 7.2), the Beta and Normal are equally good at fitting the data; however, the Gamma looks superior again as it did for SMF.

As for SMF, the TS figures (0.430 to 0.486) are comparable to those found in the North Pacific ocean experiments referenced in Chapter I. For TDF, these values show no significant difference between Gamma and Normal, or between Normal and Beta, but indicate Gamma is significantly better than Beta.

The PC, PD and PR scores also show Gamma to be the best distribution for the Fog/No Fog forecasts, but only PD shows it being significantly better than the others. For FR, all three values have insignificant differences between them, while for FA, Gamma is significantly worse than the others.

D. ENTRAINMENT (ENT)

The empirical densities (Fig. 4.2) are quite different, with the Fog population skewed to the left while the No Fog population is more evenly dispersed across the spectrum. Accordingly, there's good visual separation except for values in the vicinity of 8.

Looking at the histograms with the distribution fits (Fig. 7.3), Gamma appears to fit the Fog population best, while Beta does an excellent job with the No Fog population. For this population, the Normal is the second best visual fit, since the Gamma seems to be forcing substantial skewness where little exists.

The TS scores range from 0.346 to 0.403, somewhat less than for SMF and TDF. The Normal is significantly better than the others, perhaps a reflection of its worth as the second best fit for both populations. Normal is also significantly better than the others in the PD score.

For PR, both Normal and Gamma are significantly better than Beta, while for FA, Beta and Gamma are significantly better than Normal. For FR and PC, there are no significant differences between the three distributions.

E. LONG-WAVE RADIATION (LWR)

The empirical densities (Fig. 4.2) indicate both populations are bell-shaped with good separation.

Looking at the histograms with the distribution fits (Fig. 7.4), all three distributions fit the data quite well in both populations. The Gamma and Beta capture the slight skewness in each population, with the Gamma best able to focus on the central peak.

The TS scores range from 0.291 to 0.315, considerably lower than for SMF and TDF. There were no significant differences in the TS values; similarly for PC and FR.

The Gamma is significantly better than the others in PR, while the Normal is significantly worse than the others in FA. For PD, the Normal is significantly better than the Beta.

F. SENSIBLE HEAT FLUX (SHF)

The empirical densities (Fig. 4.2) indicate both populations are bell-shaped but with little separation.

Looking at Fig. 7.5, the Gamma does a slightly better job of capturing the peak of each population; otherwise, there's little difference between the three distributions. However, none of the distributions captures the data peaks very well.

The TS values range from 0.283 to 0.314 with no significant difference between the three distributions. The rest of the scores also show no significant difference, except for FA, where the Gamma is better than the Beta, and the PR, where it is better than the Normal.

G. STRATUS FREQUENCY (STF)

The empirical densities (Fig. 4.2) show the maximum densities of each population to be around zero, more so for No Fog than for Fog.

From the histograms (Fig. 7.6), it is difficult to tell how good the separation is likely to be. The Beta does well in capturing the high relative frequencies around zero, with the Gamma doing less well. By contrast, the Normal is clearly a poorer fit to data like these, whose distribution resembles an exponential form.

The TS values range from 0.304 to 0.467, with both Beta and Gamma doing very well compared to other reported values elsewhere. The Normal's poor fit to the data is reflected in a rather low TS value.

Except for FR, Gamma is clearly the best forecaster, with Beta and Normal ranked after it in that order. The TS, PD, FA and PR scores each show significant differences between all three distributions. In particular, it is of interest that STF is the only predictor examined whose PR scores show significant differences between all three distributions, with a rather large range of scores (0.105 to 0.226).

STF is also the only predictor whose PC score showed any significant difference between distributions; in this case, Normal was significantly worse than the other two. Only FR shows no significant difference between any of the distributions.

VIII. *CONCLUSIONS AND RECOMMENDATIONS*

A. CONCLUSIONS

This study confirms that knowledge of the underlying distributions of Fog and No Fog populations is important to the success of forecasting for these categories. In particular, it indicates that predictors whose distributions are significantly skewed definitely will be better described by a Beta or Gamma function than by a Normal. The best example of such a predictor is the Stratus Frequency, whose likelihood values peak in the vicinity of 0 and then decrease sharply toward higher values. The Normal distribution does a poor job of representing this predictor, while the Beta and Gamma do quite well.

For predictors whose distributions are roughly bell-shaped (all except STF), the results are less clear-cut, and their interpretation depends on which scoring system is being used. In general, there are fewer significant differences between the three distributions than in the case of STF. These are outlined below.

If the Threat Score is considered the single most important index of forecasting skill, this study indicates there is no significant difference between the Beta and Normal distributions for the predictors SMF, TDF, LWR and SHF. The Normal is significantly better than the Beta for ENT, while for STF, the reverse is true. The Gamma has a significantly better TS than the other two for STF, while for TDF, it is significantly better than Beta only. Otherwise, the Gamma is comparable to one and/or other of the other two.

For the Power of Detection, frequently considered a leading index of forecasting skill, there is no significant difference between Beta and Normal in the predictors TDF and SHF, while a significant difference exists for each of the other predictors. For these, the difference favors Beta only for STF, while it favors Normal for SMF, ENT and LWR. Gamma is significantly better than the others for TDF and STF and is significantly worse than them for SMF. For the other predictors, Gamma is not significantly different from one or both of the other two distributions.

While there are indications that Beta is generally competitive with the Normal, there is a number of caveats to be applied:

1. Since no forecasts were made using the "best fit" distributions of Fog and No Fog (as explained in Chapter VII), it is not clear how different the results of forecasting with such a combination of distributions would be from those shown here. In theory, such forecasts should do better.

2. The Maximum Likelihood method was not tried in this study; it would be interesting to compare the results of using this method with those used here, which employed the Method of Moments. In particular, mixing the methods between populations may slightly enhance the forecast skill.

3. Since no single distribution was clearly superior for all predictors, it is problematical how to interpret the results. There is, as yet, no universally-accepted "best" scoring system.

Except for STF, then, no definite conclusions can be reached on the basis of this study as to how much the goodness-of-fit of a distribution influences its ability to forecast. However, there is no conclusive evidence that Beta could not serve as a proxy for the Normal, subject to the comments and caveats above. Indeed, the same could be said for the Gamma, and it could well be asked whether this distribution might not do just as well as the Beta as a proxy for the Normal.

B. RECOMMENDATIONS

As a result of the work done in this study, the following ideas are offered to others doing future research in this general area:

1. Other candidate distributions, such as the T-distribution, Weibull and Lognormal, should be added to the list of distributions. Some of these may be better able to capture the unique shapes of some predictor populations.
2. The "best fit" combination of distributions, mentioned earlier, should definitely be established for each predictor population and forecasts made accordingly.
3. The Method of Moments and the Maximum Likelihood method should be examined thoroughly to see which one (if any) is best suited to fitting a given predictor. This can be done efficiently using Grafstat.
4. A means of determining which distribution is the "best fit" for a given predictor population should be definitely established. Grafstat gives a number of different "best fit" indices; it would be useful to establish if any one of these is best suited to numerical model predictor data.
5. A "best" scoring method should be defined for two-event forecasting, such as the forecasting done in this study. In particular, the PR score might be a suitable choice. In this study, PR was consistent in that, in five of the six cases, it showed Gamma as the best forecasting distribution, which reflected Gamma's generally superior way of representing the data visually on the histograms.
6. Other candidate predictors should be investigated, especially derived predictors, with a view to improving on the above results.
7. More research on the relative importance of statistical separation of the populations and the goodness of a distributional fit should be performed.

APPENDIX A

A NOTE ON GRAFSTAT

Grafstat is a product of the International Business Machine Company (IBM), and was being tested out at the Naval Postgraduate School at the time of this study. If successful, it will eventually be marketed commercially.

Grafstat is an APL (A Programming Language, used for statistics) system designed for interactive data plotting, data analysis, applied statistics, and customized graphics output. It has a full-screen, menu-driven interface, and contains a wide variety of graphics functions, a set of commonly-applied data analysis and statistical procedures, and utilities for cataloging full-screen responses, applications functions and data.

Minimal familiarity with APL is needed to use Grafstat. Users with more extensive APL backgrounds can use APL expressions as Grafstat entries, as well as integrate interactively-developed full-screen responses in the user's own APL functions.

For this study, the most useful feature of Grafstat is its ability to evaluate a set of data, fit a particular probability distribution to it, estimate the parameters of the distribution, and calculate the goodness-of-fit statistics for that distribution. Any of 18 distributions may be specified and the parameters of a distribution may be either specified or estimated.

Another feature used is the plot of a data set as a continuous curve, rather than as a histogram. This facilitated the rapid evaluation of each data set and reduced the guesswork involved in determining which candidate distributions could be ignored from the start.

APPENDIX B

VERIFICATION DATA USED IN THIS STUDY

1. The raw verification data set consists of surface weather reports for the area and time of interest, made available by the National Climatic Data Center (NCDC). It was distilled into a refined verification data set containing only those observations for which the presence or absence of fog could be definitely established. The steps in this process are as follows:

a. Number of observations received from NCDC: 12378

b. Less those deleted due to either inconsistent locations or multiple observations at the same time and place: 11303 observations remain.

c. Less those containing no definite evidence of the presence or absence of fog, as defined in paragraph 2 below: 9551 observations remain.

d. Less those for which there were no model output parameters available for that time and location: 7945 observations remain.

e. Less those located adjacent to the coastline: 5136 observations remain.

f. These 5136 observations constitute the working data set.

2. The refined data set was then divided into two categories, Fog and No Fog. Each observation was placed in the No Fog category unless it met one of the following criteria, in that order:

a. Fog was reported in Present Weather (codes 10, 11, 12, 28 and 40 through 49).

b. Fog was reported in past weather (code 4).

c. Visibility was less than 10 kilometers (ship synoptic codes 90 through 96), except when (1) winds exceeded 30 knots or (2) blowing phenomena were reported (codes 30 through 39) or (3) haze or dust was reported (codes 4 and 6) or (4) any form of moderate, heavy or frozen precipitation was reported (all codes greater than 59 other than codes 60, 61, 66, 80, and 91).

The final breakdown was as follows: Fog: 1788; No Fog: 3348.

3. Cases where visibility was 9 kilometers or less but no weather or obstruction to vision were reported presented a special problem. It was decided to accept the visibility as "truth", and to assume an obstruction to vision was present. Shipboard observation skills vary widely, and it was felt that fog was present in a large enough proportion of these cases to justify retaining these reports en masse. The omission of an obstruction to vision in such cases could be due to an incomplete observation by the observer or to a transmission problem. In any case, the presence of restricted visibility was deemed a positive indicator of fog (subject to the above exceptions), unless there was explicit evidence to the contrary.

APPENDIX C
PREDICTORS

A. NOGAPS Output Predictors

SMF..........Surface moisture flux

ENT..........Entrainment at top of marine boundary-layer

SHF..........Sensible heat flux

THF..........Total heat flux

SRA..........Solar radiation at surface

STF..........Percentage frequency of Stratus

D1000........1000 mb geopotential D-value

D925.........925 mb geopotential D-value

D850.........850 mb geopotential D-value

D700.........700 mb geopotential D-value

D500.........500 mb geopotential D-value

D400.........400 mb geopotential D-value

D300.........300 mb geopotential D-value

D250.........250 mb geopotential D-value

SST..........Sea-surface temperature

TAIR.........Surface air temperature

T1000........1000 mb temperature

T925.........925 mb temperature

T700.........700 mb temperature

T500.........500 mb temperature

T400.........400 mb temperature

T300.........300 mb temperature

T250.........250 mb temperature

EAIR.........Surface vapor pressure

E1000........1000 mb vapor pressure

E925.........925 mb vapor pressure

E850.........850 mb vapor pressure

E700.........700 mb vapor pressure

E500.........500 mb vapor pressure

UBLW.........Boundary layer zonal wind component

U1000........1000 mb zonal wind component

U925.........925 mb zonal wind component

U850.........850 mb zonal wind component

U700.........700 mb zonal wind component

U500.........500 mb zonal wind component

U400.........400 mb zonal wind component

U300.........300 mb zonal wind component

U250.........250 mb zonal wind component

VBLW.........Boundary layer meridional wind component

V1000........1000 mb meridional wind component

V925.........925 mb meridional wind component

V850.........850 mb meridional wind component

V700.........700 mb meridional wind component

V500.........500 mb meridional wind component

V400.........400 mb meridional wind component

V300.........300 mb meridional wind component

V250.........250 mb meridional wind component

VOR925.......925 mb vorticity

VOR500.......500 mb vorticity

PS...........Surface pressure

PBLD.........Planetary boundary-layer depth

STRTTH.......Stratus thickness

DRAG.........Surface drag coefficient

B. Derived Predictors

LWR..........Long-wave radiation

TDF..........Difference between 925mb air temperature and SST

AST..........Difference between surface air temperature and SST

SRH..........Surface relative humidity

TAD..........Surface air temperature advection

SAD..........Sea-surface temperature advection

APPENDIX D
FIGURES

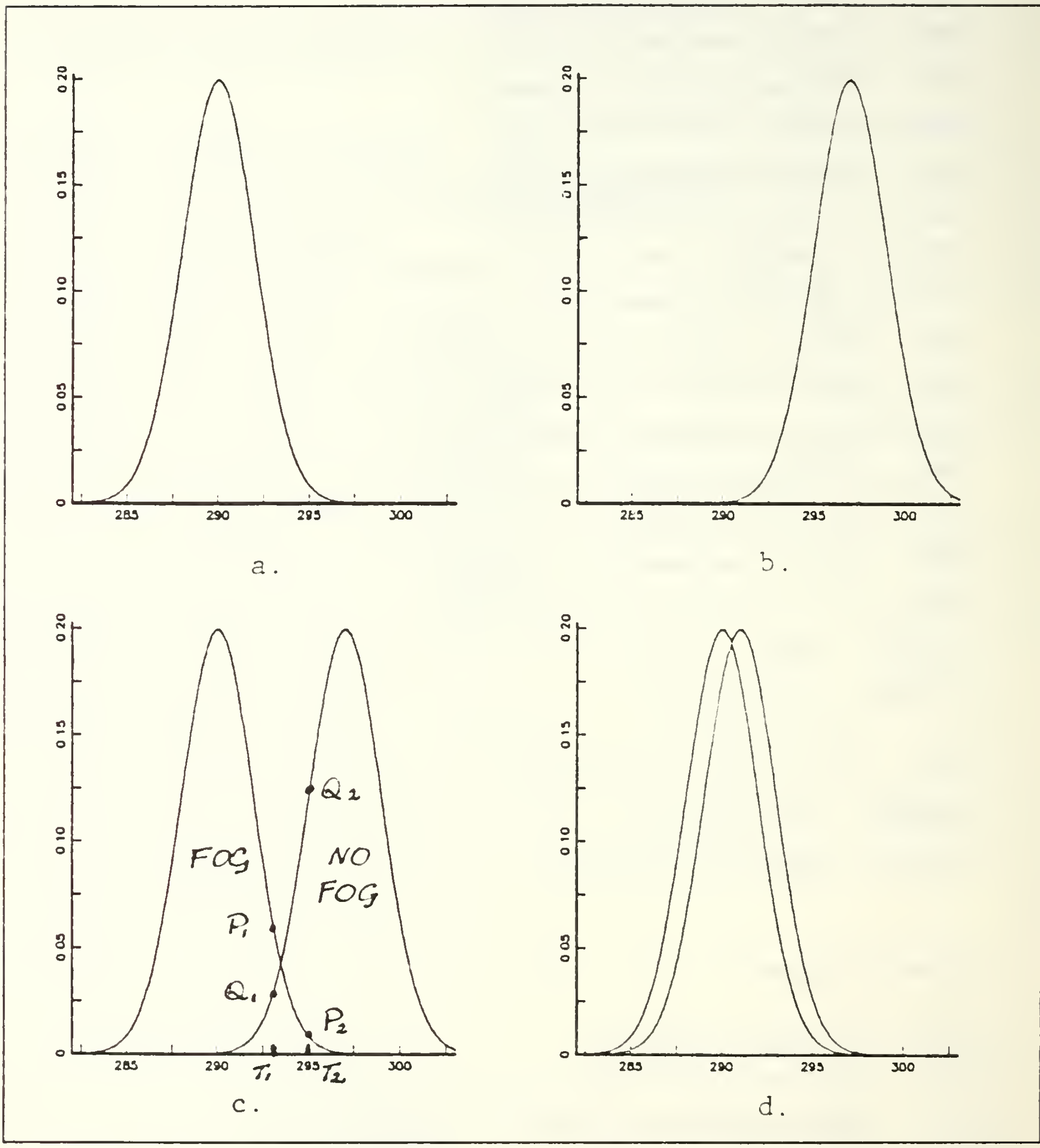

Fig. 3.1 Predictor A Distributions, Fog/No Fog.

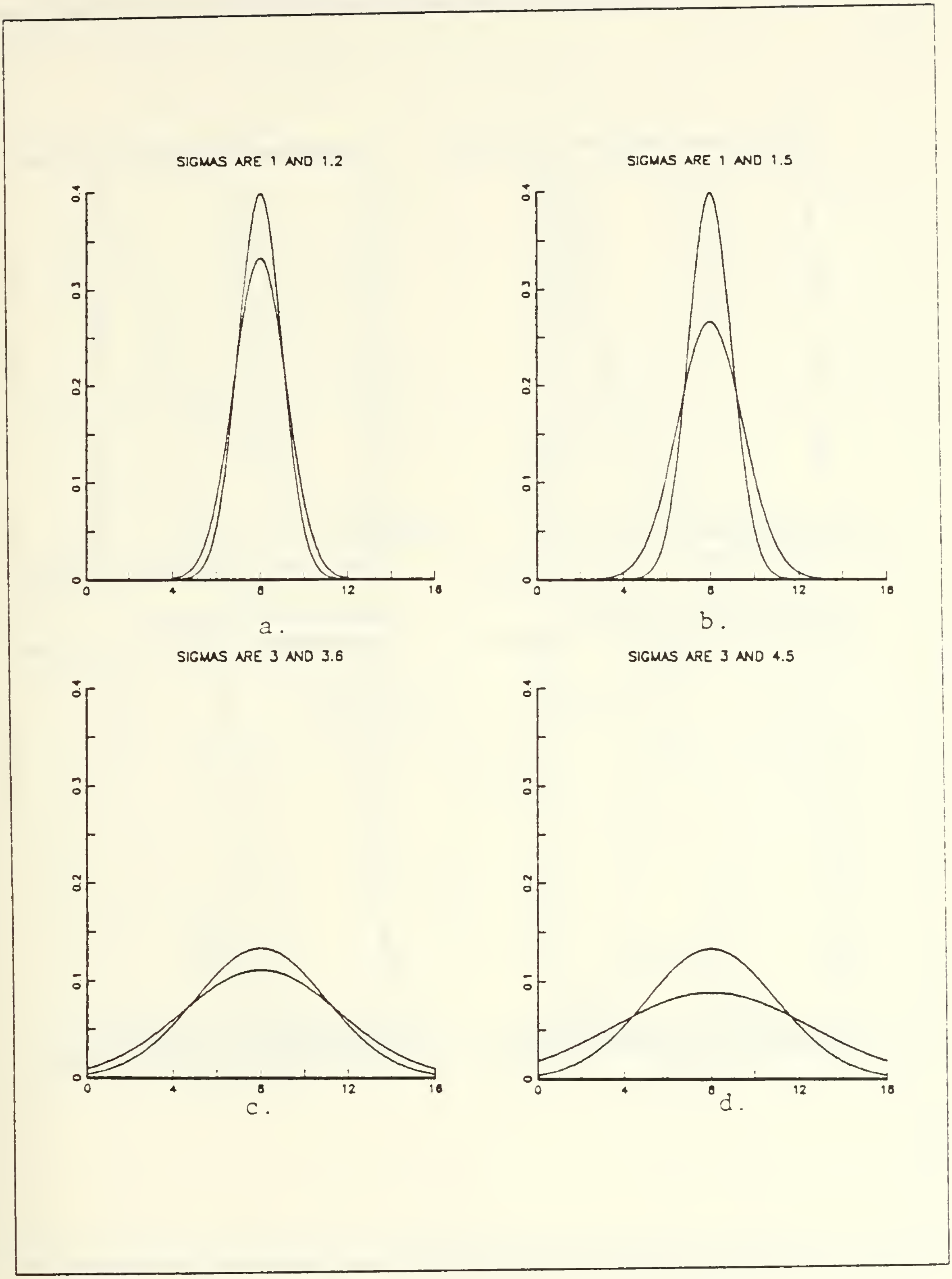

Fig. 3.2 Common Means, Different Sigmas.

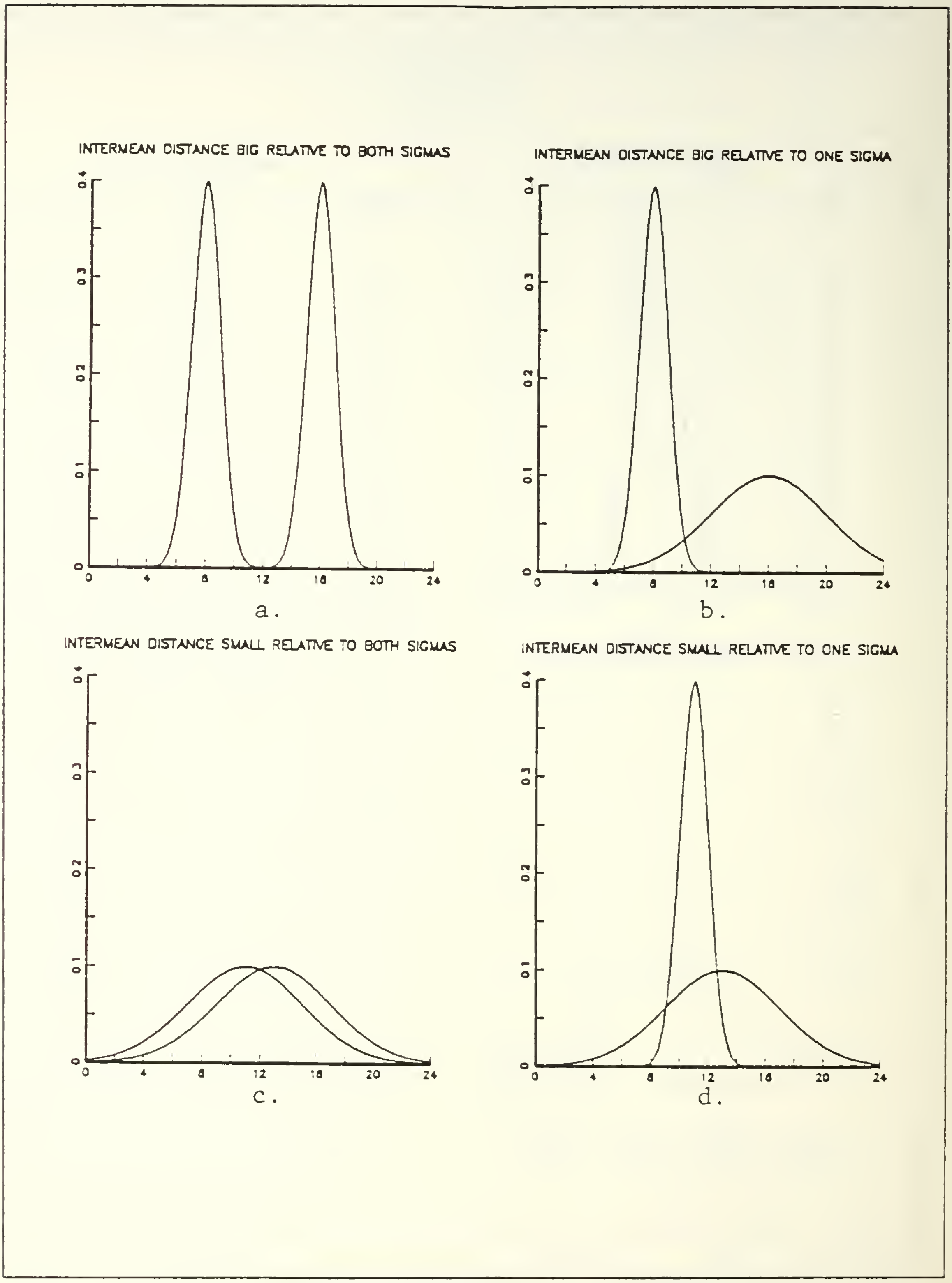

Fig. 3.3 Variations in Intermean Distance.

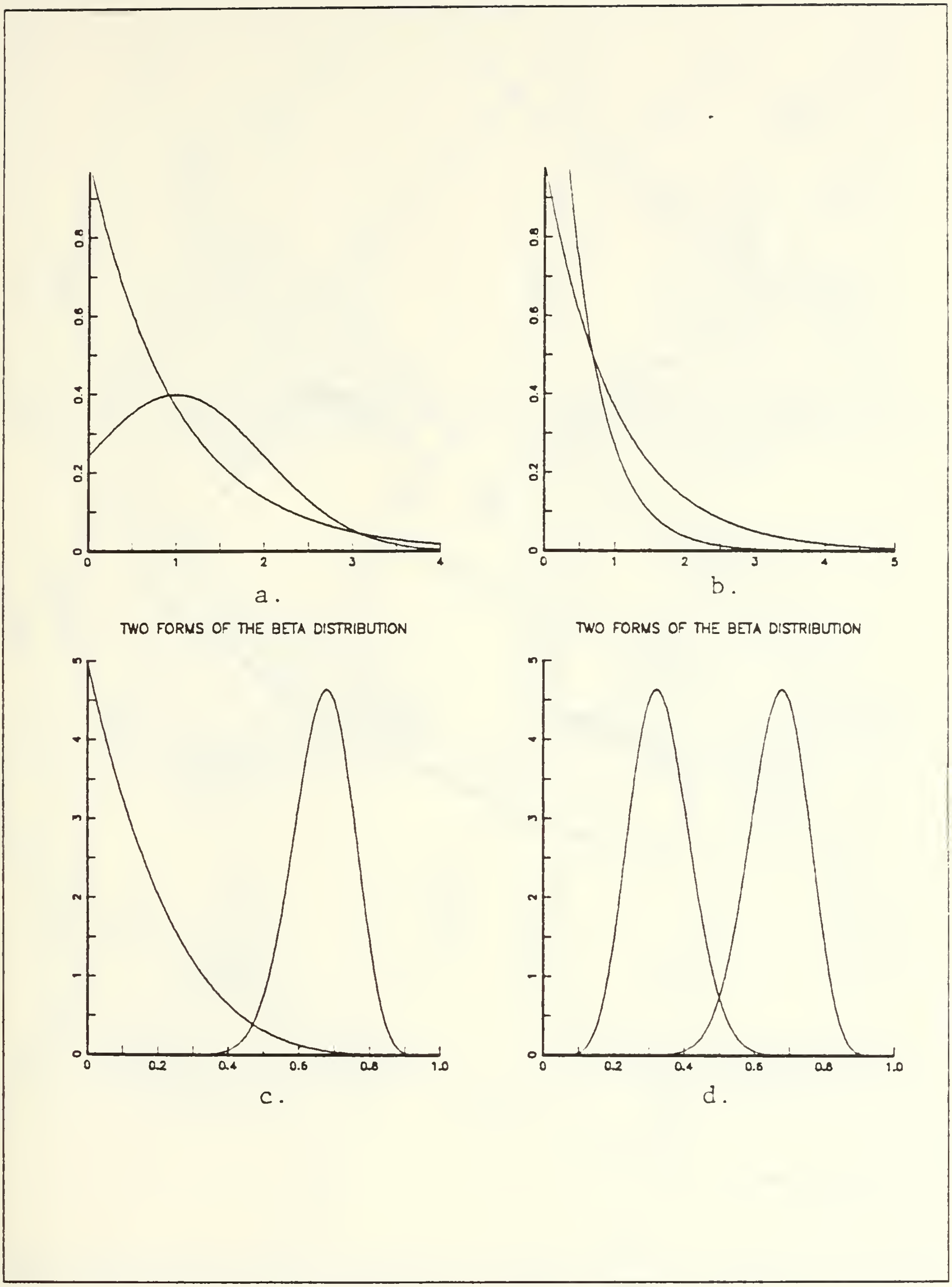

Fig. 3.4 Normal, Exponential and Beta Distributions.

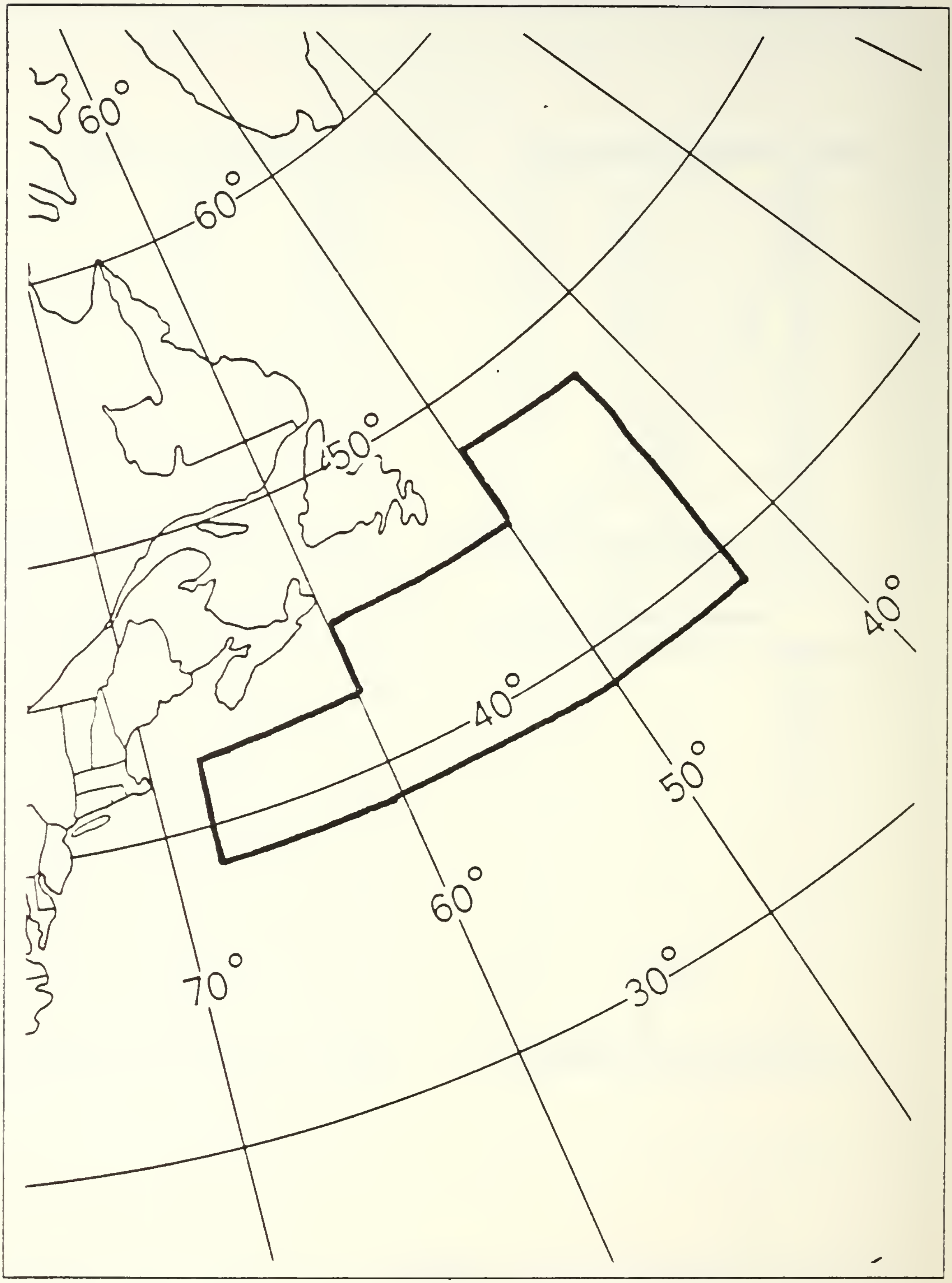

Fig. 4.1 Area of Study.

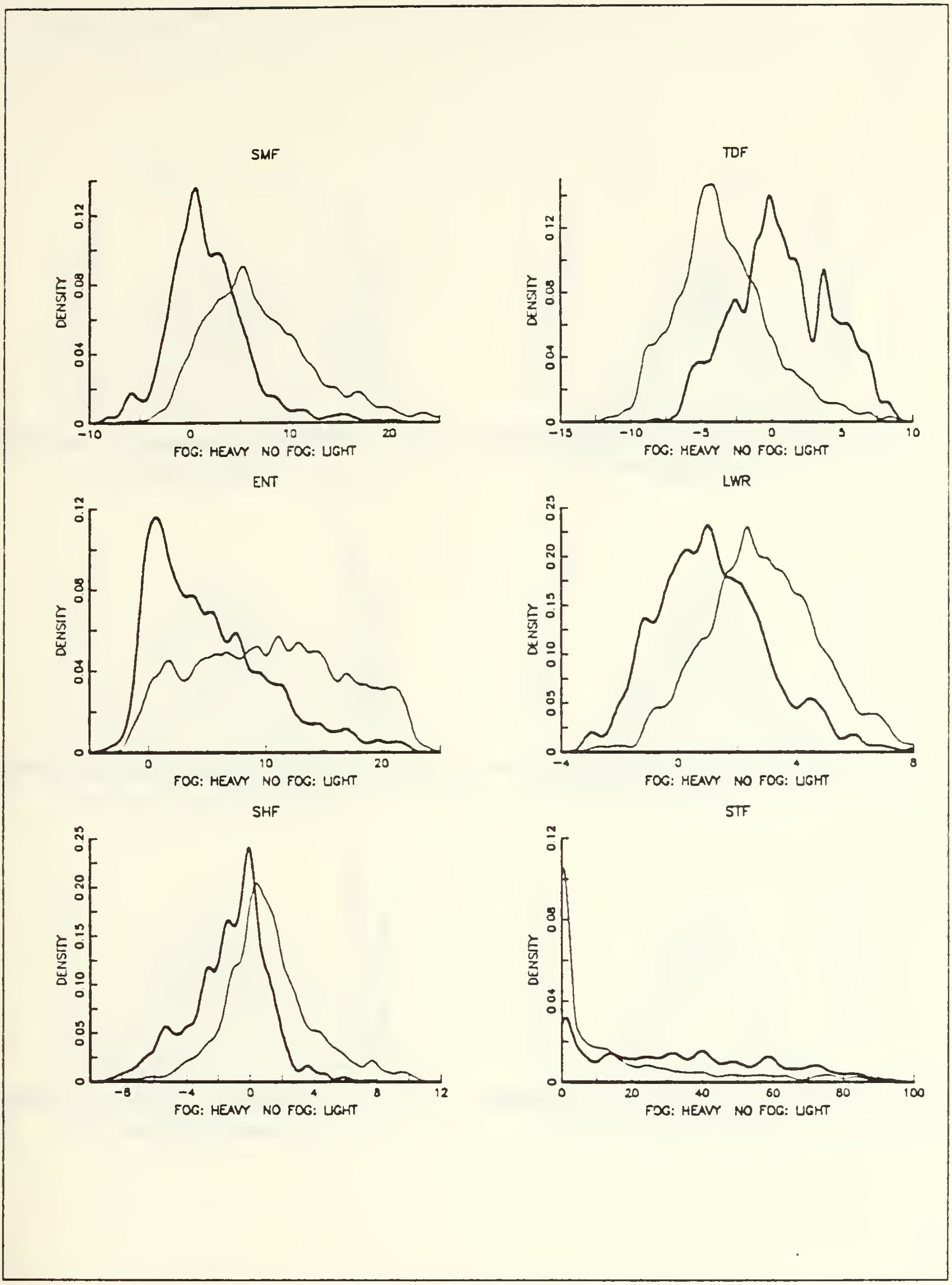

Fig. 4.2 Empirical Distribution of each Predictor.

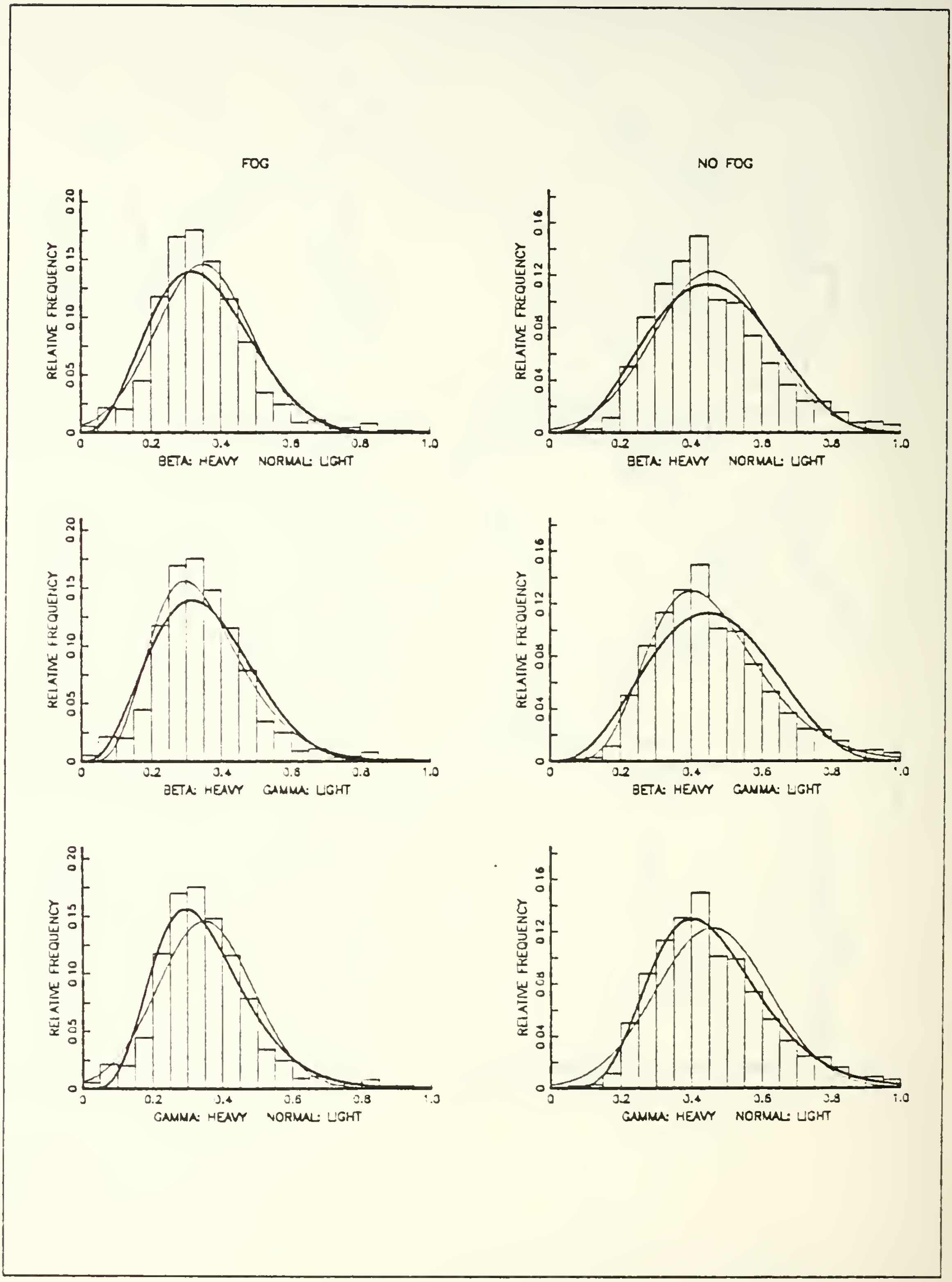

Fig. 7.1 SMF Histograms with Fitted Distributions.

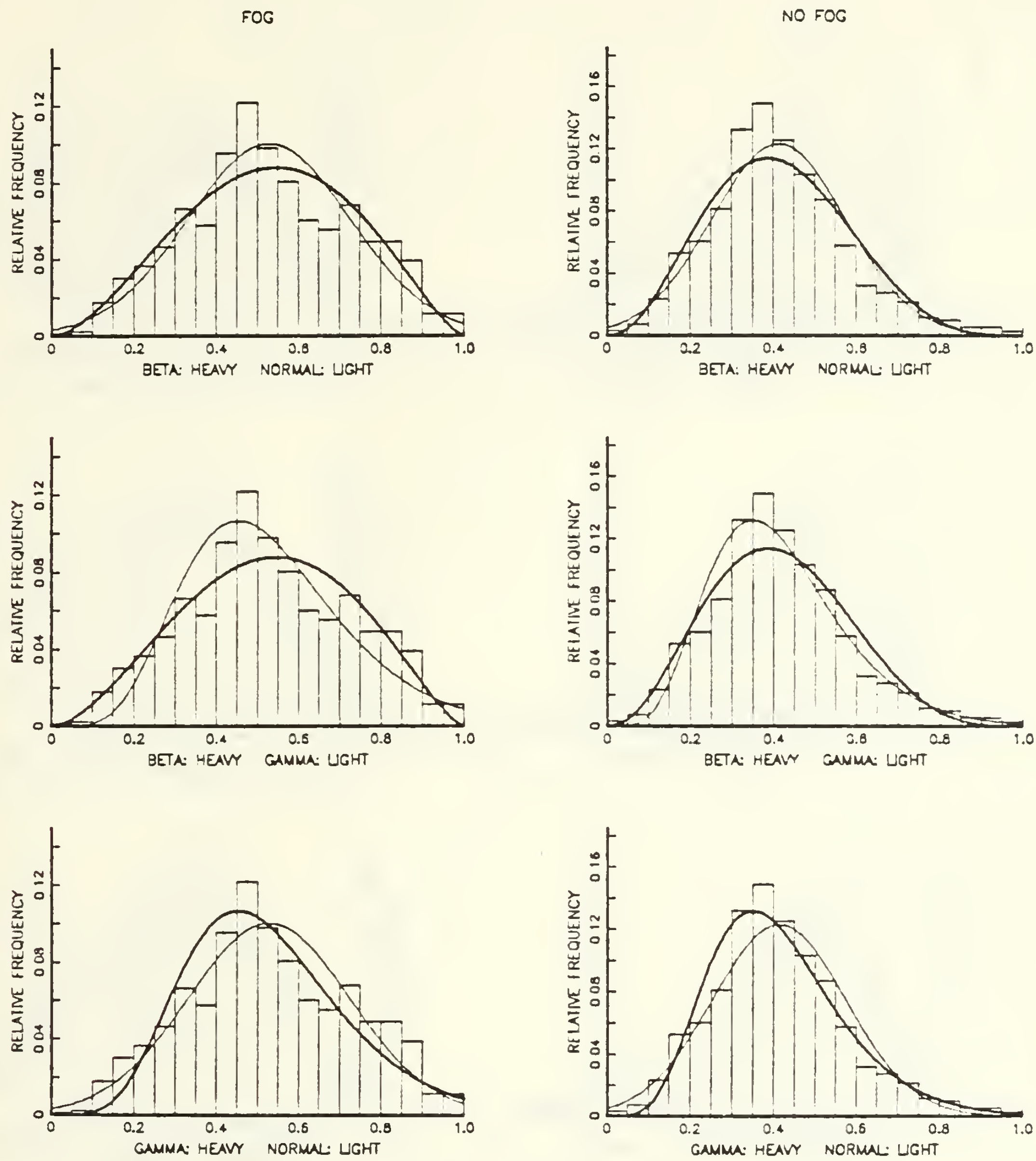

Fig. 7.2 TDF Histograms with Fitted Distributions.

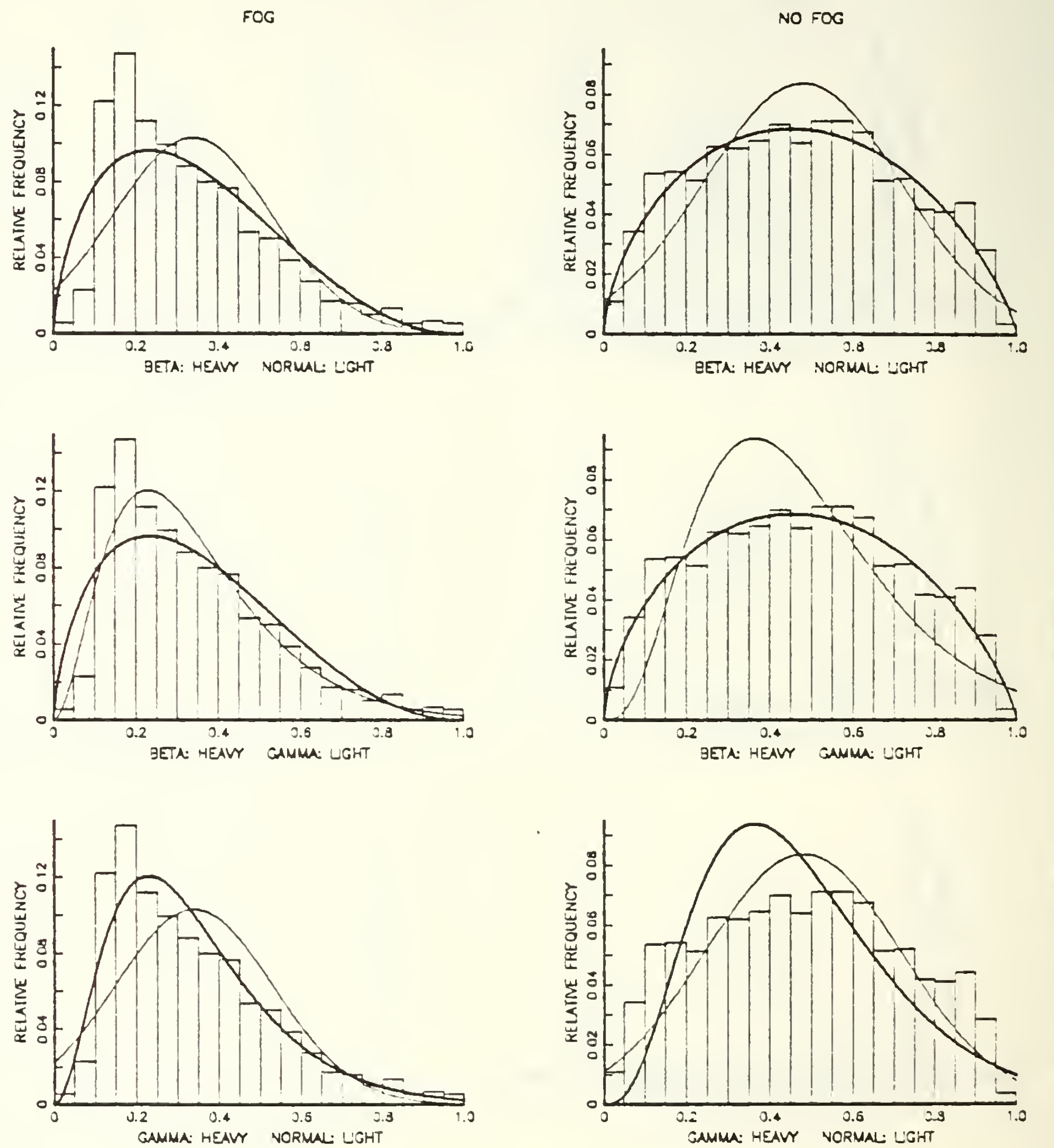

Fig. 7.3 ENT Histograms with Fitted Distributions.

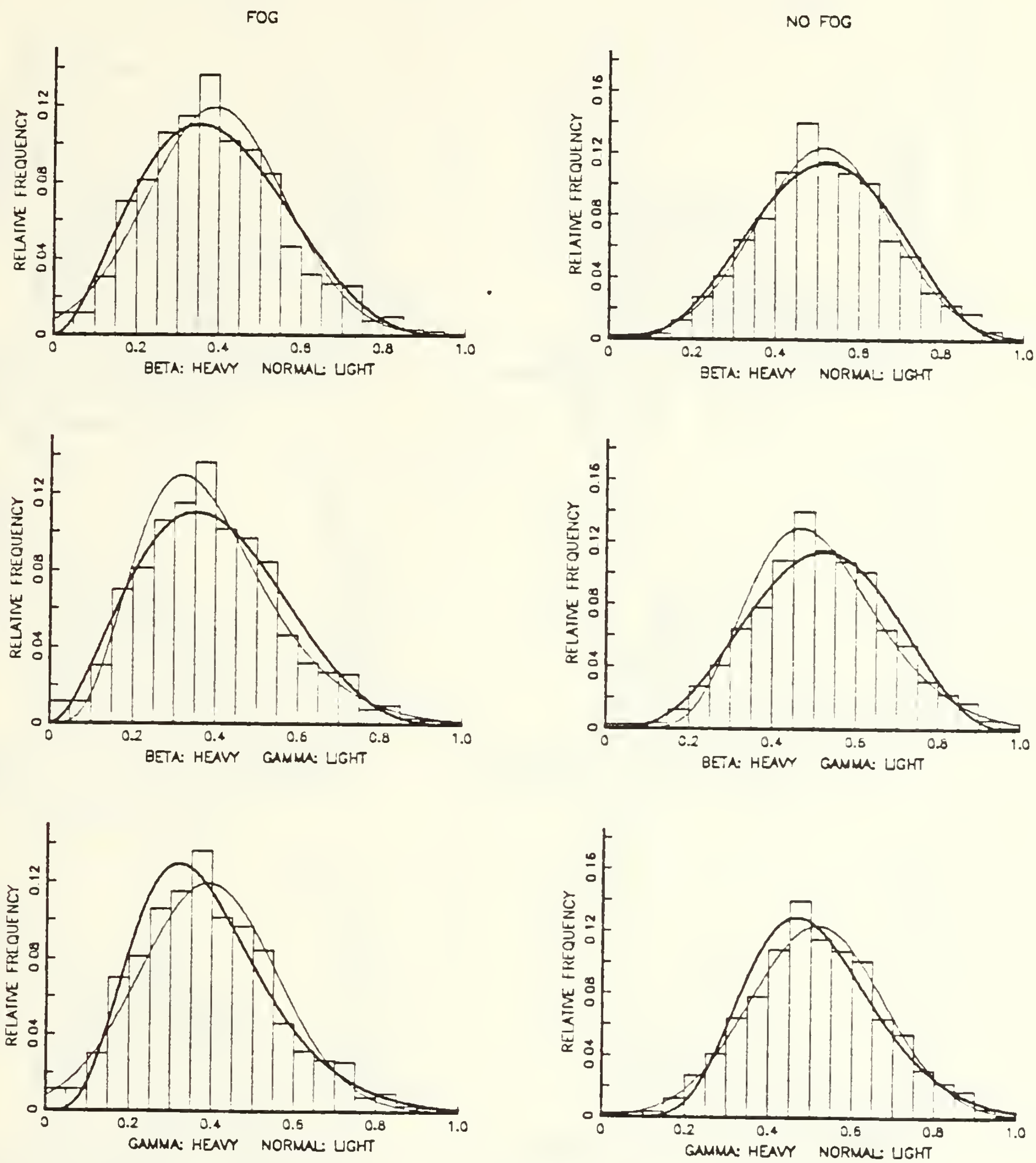

Fig. 7.4 LWR Histograms with Fitted Distributions.

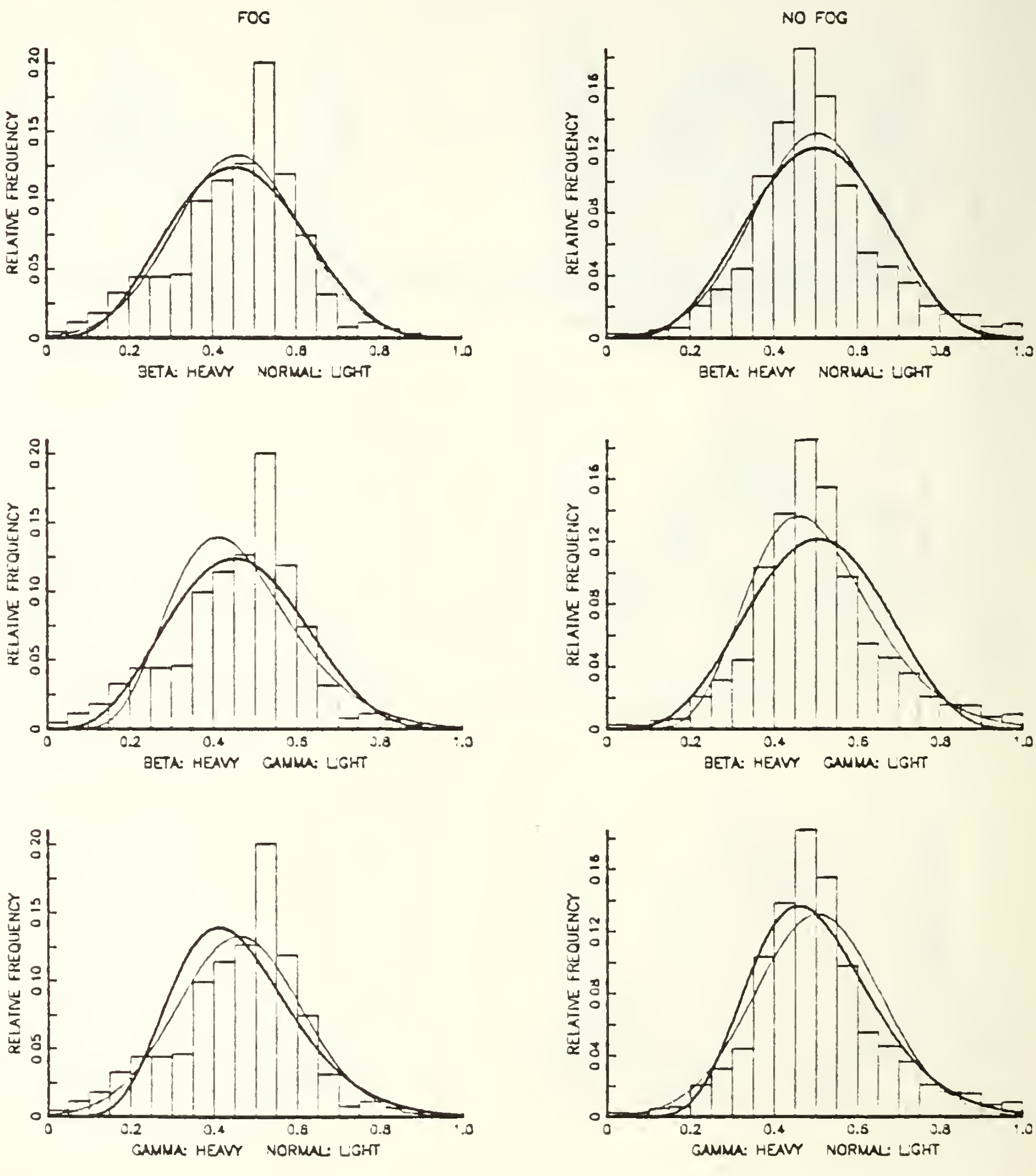

Fig. 7.5 SHF Histograms with Fitted Distributions.

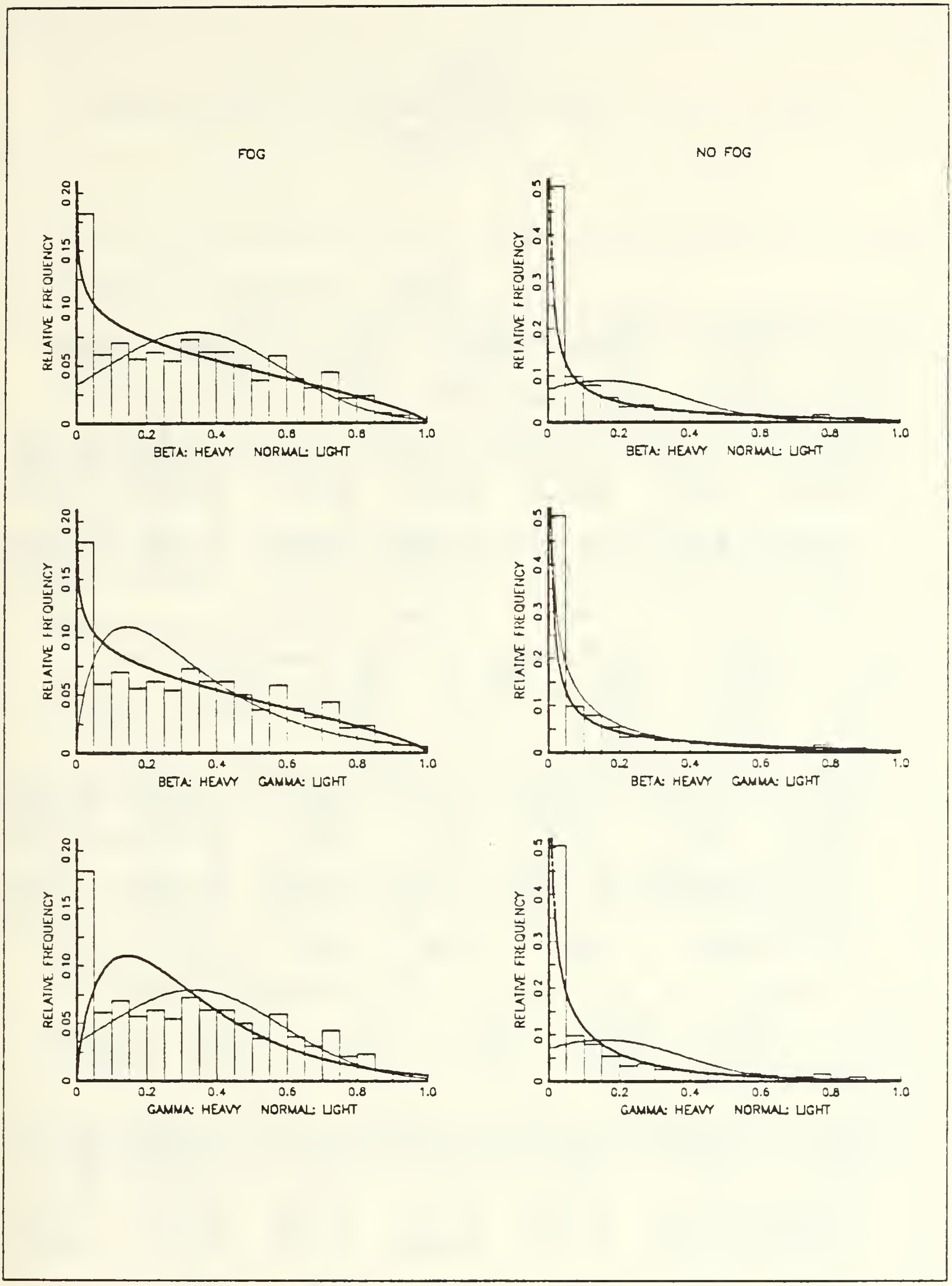

Fig. 7.6 STF Histograms with Fitted Distributions.

APPENDIX E

TABLES

TABLE 1

SCORES AND ANOVA COMPARISONS FOR SMF, TDF AND ENT

Surface Moisture Flux (SMF)

	TS	PC	PD	FR	FA	PR
Beta	0.463	0.748	0.613	0.652	0.179	0.172
Normal	0.492	0.742	0.706	0.618	0.239	0.186
Gamma	0.426	0.742	0.542	0.667	0.148	0.196
AnovaBN	0.107	0.616	0.000	0.055	0.000	0.108
AnovaBG	0.012	0.585	0.000	0.437	0.006	0.003
AnovaGN	0.002	0.841	0.000	0.005	0.000	0.232
Ranking	NB-G	BNG	N-B-G	G-B-N	G-B-N	*GNB

* Significant difference between Gamma and Beta

T_{925} - SST (TDF)

	TS	PC	PD	FR	FA	PR
Beta	0.430	0.756	0.530	0.695	0.125	0.255
Normal	0.449	0.760	0.562	0.691	0.135	0.269
Gamma	0.486	0.762	0.647	0.663	0.177	0.272
AnovaBN	0.501	0.713	0.357	0.792	0.562	0.335
AnovaBG	0.038	0.667	0.001	0.303	0.011	0.188
AnovaGN	0.173	0.802	0.009	0.369	0.042	0.721
Ranking	*GNB	GNB	G-NB	BGN	BN-G	GNB

* Significant difference between Gamma and Beta

Entrainment (ENT)

	TS	PC	PD	FR	FA	PR
Beta	0.346	0.695	0.466	0.573	0.184	0.132
Normal	0.403	0.704	0.578	0.571	0.230	0.157
Gamma	0.352	0.692	0.478	0.574	0.188	0.151
AnovaBN	0.000	0.385	0.000	0.795	0.001	0.009
AnovaBG	0.616	0.795	0.517	0.829	0.684	0.025
AnovaGN	0.001	0.443	0.000	0.778	0.004	0.493
Ranking	N-GB	NBG	N-GB	GBN	BG-N	NG-B

TABLE 2

SCORES AND ANOVA COMPARISONS FOR LWR, SHF AND STF

Long-wave Radiation (LWR)

	TS	PC	PD	FR	FA	PR
Beta	0.291	0.705	0.347	0.644	0.103	0.100
Normal	0.315	0.701	0.394	0.610	0.135	0.100
Gamma	0.293	0.701	0.355	0.626	0.113	0.126
AnovaBN	0.130	0.617	0.015	0.084	0.000	0.818
AnovaBG	0.798	0.576	0.652	0.333	0.139	0.007
AnovaGN	0.175	0.841	0.055	0.376	0.000	0.005
Ranking	NGB	BGN	*NGB	BGN	BG-N	G-BN

* Normal is significantly better than Beta

Sensible Heat Flux (SHF)

	TS	PC	PD	FR	FA	PR
Beta	0.298	0.696	0.369	0.606	0.128	0.128
Normal	0.314	0.698	0.395	0.601	0.140	0.131
Gamma	0.283	0.694	0.345	0.607	0.120	0.145
AnovaBN	0.498	0.783	0.361	0.747	0.133	0.653
AnovaBG	0.448	0.718	0.343	0.841	0.232	0.043
AnovaGN	0.148	0.643	0.059	0.732	0.006	0.100
Ranking	NBG	NBG	NBG	GBN	*GBN	**GNB

* Gamma is significantly better than Normal
** Gamma is significantly better than Beta

Stratus Frequency (STF)

	TS	PC	PD	FR	FA	PR
Beta	0.446	0.719	0.647	0.588	0.243	0.142
Normal	0.304	0.691	0.387	0.589	0.145	0.105
Gamma	0.467	0.719	0.706	0.580	0.274	0.226
AnovaBN	0.000	0.003	0.000	0.841	0.000	0.000
AnovaBG	0.150	0.823	0.004	0.458	0.001	0.000
AnovaGN	0.000	0.000	0.000	0.525	0.000	0.000
Ranking	G-B-N	GB-N	G-B-N	NBG	G-B-N	G-B-N

LIST OF REFERENCES

Barker, E.H., 1975: A Maritime Boundary Layer Model for the Prediction of Fog. *Technical Paper 4-75*, Naval Environmental Prediction Research Facility, Monterey, CA, 51pp.

Diunizio, M., 1984: *An Evaluation of Discretized Conditional Probability and Linear Regression Threshold Techniques in Model Output Statistics Forecasting of Visibility over the North Atlantic Ocean.* Master of Science Thesis (R. J. Renard, advisor), Department of Meteorology, Naval Postgraduate School, Monterey, CA, 233 pp.

Elias, K.C., 1985: *Forecasting Atmospheric Visibility over the Summer North Atlantic using the Principal Discriminant Method.* Master of Science Thesis (R. J. Renard, advisor), Department of Meteorology, Naval Postgraduate School, Monterey, CA, 112 pp.

Fatjo, S.J., 1986: *A Study to Determine the Relative Skill of Four Model Output Statistics Prediction Methods using Simulated Data Fields.* Master of Science Thesis (R. J. Renard, advisor), Department of Meteorology, Naval Postgraduate School, Monterey, CA, 72 pp.

Glahn, H.R., 1983: MOS Support for Military Locations from the Technique Development Laboratory. *TDL Office Note 83-9*, 10 pp.

Glahn, H.R. and D.A. Lowry, 1972: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. *J. Appl. Meteor.*, 11, pp. 1203-1211.

Karl, M.L., 1984: *Experiments in Forecasting Atmospheric Marine Horizontal Visibility using Model Output Statistics with Conditional Probabilities of Discretized Parameters.* Master of Science Thesis (R. J. Renard, advisor), Department of Meteorology, Naval Postgraduate School, Monterey, CA, 165 pp.

Koziara, M.C., R.J. Renard and W.J.Thompson, 1983: Estimating Marine Fog Probability using a Model Output Statistics Scheme. *Monthly Weather Review*, 111, pp. 2333-2340.

Mack, E.J., C.W. Rogers and B.J. Wattle, 1983: An Evaluation of Marine Fog Forecasts and a Preliminary Design for a Marine Obscuration Forecast System *Calspan Report Number 6866-M-1*, Advanced Technology Center, Buffalo, NY, 250 pp.

Pilie, R.J., E.E. Mack, C.W. Rogers, U. Katz and W.C. Kochmond, 1979: The Formation of Marine Fog and the Development of Fog-Stratus Systems along the California Coast. *Journal of Applied Meteorology*, Volume 18, Number 10, October 1979, pp. 1275-1286.

Renard, R.J., 1976: The Observation, Analysis, Forecasting and Climatology of Marine Fog. *WMO Geneva, Switzerland, Paper 454*, pp. 211-223.

Renard, R.J., and W.J. Thompson, 1984: Estimating Visibility over the North Pacific Ocean using Model Output Statistics. *National Weather Digest*, Vol 9, Number 2, pp. 18-25.

Renard, R.J., R.E. Englebretson and J.S. Daughenbaugh, 1975: *Climatological Marine-Fog Frequencies Derived from a Synthesis of Visibility-Weather Group Elements of the Transient Ship Synoptic Reports*, NPS Report 51Rd75041, Naval Postgraduate School, Monterey, CA, 47 pp.

Wheeler, S.E and D.F. Leipper, 1974: *Marine Fog Impact on Naval Operations*, NPS Report 58Wh74091, Department of Oceanography, Naval Postgraduate School, Monterey, CA, 118 pp.

Willms, G.P., 1975: *A Climatology of Marine-Fog Frequency for the North Pacific Ocean Summer Fog Season*, Master of Science Thesis, Department of Meteorology, Naval Postgraduate School, Monterey, CA, 59 pp.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, CA 93943-5000	2
3.	Chairman, Code 63Rd Department of Meteorology Naval Postgraduate School Monterey, CA 93943-5000	7
4.	Chairman, Code 68Mr Department of Oceanography Naval Postgraduate School Monterey, CA 93943-5000	1
5.	Mr R. Hale, Code 63 Department of Meteorology Naval Postgraduate School Monterey, CA 93943-5000	1
6.	Mr P. Lowe Naval Environmental Prediction Research Facility Monterey, CA 93943	1
7.	Captain Oliver J. Muldoon AFGL/LYA Hanscom AFB, MA 01731-5000	3
8.	Director Naval Oceanography Division Naval Observatory 34th and Massachusetts Avenue NW Washington, DC 20390	1
9.	Commander Naval Oceanography Command NSTL Station Bay St. Louis, MS 39522	1
10.	Commanding Officer Naval Oceanographic Office NSTL Station Bay St. Louis, MS 39522	1
11.	Commanding Officer Fleet Numerical Oceanography Center Monterey, CA 93943	1
12.	Commanding Officer Naval Ocean Research and Development Activity NSTL Station Bay St. Louis, MS 39522	1
13.	Commanding Officer Naval Environmental Prediction Research Facility Monterey, CA 93943	1

14. Chairman, Oceanography Department 1
U.S. Naval Academy
Annapolis, MD 21402

15. Chief of Naval Research 1
Naval Ocean Research and Development Activity
800 N. Quincy Street
Arlington, VA 22217

16. Office of Naval Research (Code 420) 1
Naval Ocean Research and Development Activity
800 N. Quincy Street
Arlington, VA 22217

17. Library 1
University College
Galway
Ireland

18. AFIT/CIRF 1
Wright-Patterson AFB
OH 45433

19. USAFETAC/TS 1
Scott AFB
IL 62225

20. AFGWC/TS 1
Offutt AFB
NE 68113

219559

Thesis
M8852 Muldoon
c.1 Relating the distributional character of numerical model output parameters to the occurence of fog over the North Atlantic Ocean.